

METODIKA STATISTICKÉHO ZPRACOVÁNÍ

Kapitola shrnuje základní metodiku statistických postupů, které obvykle bývají provedeny před zahájením zpracování dat.

Základem zpracování dat je ověření jejich kvality ve smyslu minimalizace možných nesrovnalostí, které jsou dány nepřesným, či chybným vyplněním, resp. následným kódováním dotazníku. Předpokladem kvality dat je pochopitelně kvalita samotného dotazníku. Otázky v dotazníku by proto měly být formulovány s ohledem na obecně přijímané zásady.¹

Úvodní část zpracování dat se – tak jak je obvyklé - sestávala z

- kódování dat,
- jejich editování a
- čištění.

Data získaná v šetření jsou obvykle převedena do elektronické podoby. Pro kódování dat bývá obvykle zvolen speciální software (např. program SPSS, v případě menších šetření lze vystačit i s EXCELEM), přičemž otázky by měly být kódovány v logice dotazníku. Jedna z podstatných otázek, kterou je nutno řešit již při kódování dat, je způsob přepisu chybějících (nevyplněných) odpovědí. Zásady zpracování dat jednoznačně rozlišují několik označení (dle důvodů) pro chybějící (nevyplněná data).² Odpovědi označené jako chybějící jsou poté z dalších analýz vyloučeny. U jednotlivých otázek se ve výstupech objevují různé počty odpovědí, neboť do analýz zahrnujeme pouze validní (neboli ne-chybějící) data.

Obě další etapy – editování a čištění dat – mají stejný cíl. Jejich úkolem je minimalizovat počet případných chyb ve vyplněných odpovědích. První požadavek bývá označován jako **úplnost dat**. Jedná se o minimalizaci chybějících odpovědí a naplňování zejména výše zmíněným kódováním nevyplněných odpovědí. Druhá zásada se označuje jako **konzistence** a její podstatou je požadavek, aby data splňovala základní parametry, či logické vazby. Jedná se například o ty otázky, kde je nabídnuto několik možností, přičemž výsledek by měl dávat celek (obvykle 100 %). Konzistence pochopitelně také znamená, aby odpovědi byly vyplňovány tazateli jednotně.

Chybné nebo nekonzistentní odpovědi, které vyplynou z výše uvedených dvou etap, je možno posoudit a očistit dvěma možnými způsoby

- vyloučit z primární matice,
- nahradit chybou hodnotu hodnotou správnou.

V rámci této fáze čištění dat je nutno se držet obecně přijímané zásady posvátnosti sesbíraných dat. Tato zásada ve stručnosti říká, že hodnoty uváděné v dotazníku je možno měnit jen v případech, kdy je „více než jasné“, že jsou chybné.³ Stejně tak je nutno dodržovat zásadu „doplňování dat v nejmenší možné míře“. Jinými slovy, v souladu s výše uvedenou

¹ Jedná se zejména o stručnost a jednoznačnost otázek, vyhnout se zavádějícím otázkám apod. Více viz např. IAROSSO (2006).

² Obvykle se označují jako NA (not applicable=nepoužitelná), NP (not provided=neposkytnutá), DK (don't know=nevím) a REF (refuse to answer=odmítl odpovědět). Vztah mezi těmito typy označení chybějících hodnot proměnných uvádí IAROSSO (2006).

³ IAROSSO (2006)

zásadou, vstupovat do primární matice jen v případech, kdy je nezpochybnitelné, že respondent hodlal na danou otázku odpovědět právě tímto způsobem.⁴

Výsledkem této přípravné fáze zpracování dat tedy je primární matice s konzistentními daty, přičemž všechny logické vazby mezi nimi (uváděné v dotazníku) jsou splněny. Je tedy možné data použít pro další statistické zpracování.

Typy proměnných vstupujících do prezentovaných analýz

V dotazníku, který bývá základem většiny šetření, se využívá několika typů proměnných. Kromě menšího množství otevřených otázek,⁵ obsahuje dotazník obvykle především uzavřené otázky. Uzavřené otázky respondentovi nabízejí různé (předem dané) varianty odpovědí. Na základě těchto typů odpovědí je možno rozlišit jednotlivé proměnné (které odpovědi na otázky v dotazníku zastupují) na několik typů.

Vzhledem k tomu, že různé typy proměnných je nutno v rámci jejich statistického vyhodnocování pojímat odlišně, a také k tomu, že mezi různými typy proměnných existují přímé vztahy, uvádíme nejprve jejich varianty:⁶

- **Nominální proměnné**

Slouží k pojmenování (odsud nominální) a odlišení různých typů převážně kvalitativních znaků. Hodnotami tohoto typu proměnných jsou především texty, písmena, či znaky. S výjimkou tzv. binárních⁷ nelze s nominálními proměnnými provádět žádné aritmetické operace (sčítat, odčítat, apod.)

- **Ordinální proměnné**

Jsou obdobou nominálních proměnných, tentokrát však můžeme jednotlivé hodnoty proměnné mezi sebou porovnávat na číselné ose (ve smyslu větší/menší). Lze je dále statisticky vyhodnocovat i s použitím aritmetických operací.⁸

- **Intervalové proměnné**

Obecnější variantou ordinální proměnné jsou intervalové proměnné, které lze porovnat i z hlediska jejich vzájemné velikosti (tedy o kolik je jedna větší než jiná).

- **Poměrové (podílové) proměnné**

Jsou obecnější variantou intervalových proměnných. Odlišují se tím, že umožňují i srovnání ve smyslu poměru (kolikrát je jedna proměnná větší než druhá). Poměrová proměnná má tedy (na rozdíl od intervalové) jednoznačný výchozí bod (bod nula).

Nominální a ordinální proměnné jsou také společně označovány jako kvalitativní proměnné, zatímco poměrové a intervalové jsou řazeny mezi kvantitativní proměnné.

⁴ Jak uvádí FELLEGI, HOLT (1976) cílem dotazníkových šetření je sbírání a ne vyrábění dat.

⁵ Otázky u nichž je respondent nucen dopověd' doplnit v podobě slovního vyjádření nikoli jako výběr z nabídnutých variant.

⁶ Podrobněji viz např. BABBIE, HALLEY, ZAINO (2000)

⁷ Typický příklad binární nominální proměnné je např. pohlaví, kdy proměnná může nabývat pouze dvou hodnot (muž, žena).

⁸ Otázky obvykle mají varianty odpovědí na stupnici např. 1=nedůležitá po 5=vysoce důležitá.

POUŽITÁ LITERATURA

BABBIE, E. HALLEY, F. ZAINO, J. *Adventures in social research : data analysis using SPSS for Windows 95/98*. Thousand Oaks : Pine Forge Press, 2000. 451 s. ISBN 0761986766

FELLEGI, I. P. HOLT, D. A Systematic Approach to Automatic Edit and Imputation. *Journal of the American Statistical Association*, Vol. 71, No. 353. (Mar., 1976), pp. 17-35.

IAROSSO, G. *The power of survey design : a user's guide for managing surveys, interpreting results, and influencing respondents*. Washington, D.C. : World Bank, 2006. 262 s. ISBN 0821363921